DOCUMENT RESUME

ED 206 649 TM 810 539

AUTHOR Smith, Philip L.; Kane, Michael
TITLE Multivariate Generalizability: An Application to Student Course Ratings.
PUB DATE Apr 81
NOTE 27p.: Paper presented at the Annual Meeting of the American Eudcational Research Association (65th, Los Angeles, CA, April 13-17, 1981).

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Course Evaluation; Item Analysis; *Multivariate Analysis; Post Secondary Education; Questionnaires; *Student Evaluation of Teacher Performance; *Weighted Scores
IDENTIFIERS Composite Scores; *Generalizability Theory

ABSTRACT

The aims of this study were: (1) to review and discuss alternatives for evaluating members of a set of scores which may be interpreted independently as contributing to a composite score, or as a member of a profile and (2) to examine multivariate generalizability theory, as suggested by Joe and Woodward, applied to student course/instructor ratings by analyzing course evaluation questionnaires. Described are the procedures which were applied to data obtained from seven samples on two campuses to derive multivariate generalizability coefficients associated with item score patterns that maximally discriminate between courses and instructors. Analysis results for the two basic design figurations (instructors nested within courses and courses nested within instructors), and the results of split halves analyses, used to ascertain the stability of composite weights across samples, are discussed and displayed in tabular form. Discussion of the results of the study stresses the limitation of the Joe and Woodward procedure since it was found that weights assigned to different items were highly variable across split-halves analyses. It is suggested that the stability of results across samples will be greatest when the number of variables is small or the sample size very large. (Author/AEF)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED206649

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

Multivariate Generalizability:

An Application to Student Course Ratings

Philip L. Smith

University of Wisconsin-Milwaukee

Michael Kane

National League for Nursing

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

P. L. Smith

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM 810 539

Paper presented at American Educational Research Association Annual Meeting, Los Angeles, April, 1981.

Introduction

Recently, interest has surfaced in application of multivariate generalizatility theory as suggested by Joe and Woodward (1976). In their application, Conger and Lipshitz (1973) and Webb and Shavelson (1981) illustrate both the usefulness and unique interpretation stemming from multivariate G studies. Essentially, multivariate G studies are conducted to find the maximally reliable profiles among a set of dependent variables. The method is adequately described in Joe and Woodward (1976). The purpose of this paper is twofold. First, a primary purpose of this paper is to review and discuss alternatives for evaluating members of a set of scores which may be interpreted independently, as contributing to a composite score, or as a member of a profile. How such a score is envisioned has significant bearing upon its treatment and role within the context of generalizability theory. Second, it is the purpose of this paper to illustrate multivariate G theory applied to student course/instructor ratings. It is concluded that the multivariate treatment of scores can provide information different from (and not necessarily consistent with) the univariate treatment, and that the choice of analytical method is a function of how the data are to be used in a decision (D) study context.

Components, Composites, and Profiles

In general, behavioral measures can be considered either or both multifaceted and multivariate. The conceptualization of the measure(s) with respect to each of these terms should be relatively independent. The language of generalizability theory considers a measure multifaceted when several conditions (facets) of its measurement are allowed to vary. A measure is multivariate when it results in a set of scores that are to be interpreted as

a pattern or profile, each member of which implies a somewhat different construct (Cronbach, Gleser, Nanda, & Rajaratnam, 1972). Of course, a set of scores could be considered multivariate with respect to a set of fairly narrowly defined constructs, and univariate (i.e., multiple observations of the same construct) with respect to a single broadly defined construct. Since the interpretation to be given to the scores determines the appropriate analytical strategy, the use of a dependent variable (or set of dependent variables) in decision making (D studies) must be well thought out and specified. A third option is the treatment of an observation as part of the sum of a composite score. If decisions are made on the basis of a single composite score, the distinction between univariate and multivariate analyses is less important from an analytical perspective. In this case, the score is seen as a contributing member of the construct defined by the composite, and thus contributing some unique and some common variance to this score.

As an example, consider a simple classroom observation scheme. We can rate the same behaviors on various occasions using various raters in which case the measure is multifaceted, but not multivariate. We could rate several conceptually distinct behaviors perhaps on several occasions with several raters; in this case the measure is both multifaceted and multivariate. Or a weighted sum of ratings of different aspects of the same behavior can be interpreted as a composite, in which case the measure could be considered multivariate or univariate, depending upon the construct interpretation intended.

Sometimes the distinction is not obvious and serious consideration needs to be given to how the scores are to be used before a decision is made to treat a set of scores as multivariate or as having arisen from sampled levels of a

facet. Consider, for example, the language portion of the Iowa Test of Basic Skills (ITBS). There are four language subtests that result in a language subscore. Should these tests be considered levels of a facet or distinct behaviors that result in a pattern or profile? An analagous situation holds for the subtests of the WISC-R. Essentially, it may be argued here that if the subtests are intended to have the same construct interpretation, then differences in the conditions for arriving at these scores should be considered conditions of a facet and a univariate procedure is appropriate. If, on the other hand, subtests scores are seen as distinct constructs then a multivariate or profile analysis would be appropriate. Returning to the ITBS example, if the subscores of the language portion of the ITBS were considered as separate indicators of general language achievement, then they might be characterized as levels of a facet that differ in item form, content, etc. On the other hand, if these scores are viewed as distinct variables, each of which contributes a large unique portion to overall language achievement then a profile interpretation and multivariate analysis would be more appropriate. Of course, a single score could be interpreted both ways. For instance, the information subtest of the WISC-R can and has been used both as a member of a profile and as an indicator of verbal comprehension (Kaufman, 1979).

Application to Student Instructor/Course Ratings

One area in particular where the distinction between a multivariate and univariate interpretation has not been explicitly considered is in the analysis of student course/instructor ratings. The reliability and validity of such ratings have been extensively examined using both classical measurement theory (see, for example, Doyle, 1975) and generalizability theory (see Kane, Gillmore, & Crooks, 1976; Kane and Brennan, 1977; Gillmore, Kane, &

Naccarato, 1978; Smith, 1979). Although the literature on these ratings routinely treats them as complex and multifactored, the analyses of the technical adequacy of the instruments has relied almost exclusively on univariate methods. Although some studies examine these ratings at the item level, a more common approach is to treat items as if they were all indicators of the same construct (i. e., by summing to a total or treating items as levels of a facet). This practice may be justified if the items are viewed as separate indicators of an overall construct of course/instructor effectiveness. However, such instruments are often justified for their diagnostic value by assuming that distinct information is contained in the separate items, and for such uses, a profile interpretation may make more sense.

One option for analyzing the ratings so that profiles are considered is multivariate generalizability analysis, which allows for multiple facets and multiple dependent variables. Such an analysis results in composites (and their reliabilities) that are determined by the data to be maximally reliable (see Conger & Lipshitz, 1973; Joe & Woodward, 1976). The advantage of such composites for course/instructor evaluations is that not only do they recognize the unique contribution of items to a pattern, but also reveal which item profiles contribute to maximum differentiation of courses and/or instructors. Such an analysis is particularly relevant to the analysis of course/instructor ratings given the current literature on the distinction between effects produced by instructors and effects produced by courses (see Gillmore et al., 1978; Smith, 1979). The uses of multivariate generalizability procedures to identify composites that provide maximally reliable ratings for discriminating among courses on the one hand and/or instructors on the other hand, could highlight differences among items in how they are used by students to evaluate courses and instructors respectively.

The focus of the current study is on the discrimination of courses independent of instructor, and instructor independent of courses. Two design configurations, one with students nested within courses within instructors and the other with students within instructors within courses are thus relevant. Of course, since items are to be treated not as separate observations of the same construct, but rather as scores with independent construct interpretations, items are not considered a facet in the multivariate analysis.

Method

The main focus of this paper is on the multivariate (profile) interpretations for student course/instructor ratings. Procedures suggested by Woodward & Joe (1976) were applied to several sets of data to provide multivariate generalizability coefficients associated with item score patterns that maximally discriminate between courses, on the one hand, and instructors, on the other. Details of these procedures are provided below.

Samples. In all, seven separate samples of instructor/course ratings were used for this analysis. These samples were drawn from two campuses and were configured as follows: Sample 1-Campus 1, single department, two courses nested within each of 15 instructors; Sample 2-Campus 1, single department, two instructors nested within each of 11 courses; Sample 3-Campus 2, two sections nested within each of 5 instructors, all sections of the same course; Sample 4-Campus 2, single department, two courses nested within each of 5 instructors; Sample 5-two courses nested within each of 6 instructors, each instructor from a different department; Sample 6-same department, two instructors nested within each of 5 courses; and Sample 7-same department, two courses within each of 8 instructors. The number of students nested within each of the instructor/course combinations ranged between 10 and 18, but was balanced for each sample by randomly eliminating cases.

These samples have two basic configurations: students within courses within instructors (samples 1, 3, 4, 5, 7) and students within instructors within courses (samples 2, 6). The first of these configurations is useful for studying instructor differences and the second for course differences. Within each of these configurations there are several differences among the analyses. First, both configurations involve samples from two different universities, thus permitting a comparison across institution. (As noted below, this comparison is confounded by a difference in the evaluation forms used.) Within the configuration where courses are nested within instructors, there are four different samples; there is an analysis with several sections of the same course, several courses in the same department (for two different departments), and several courses across campus.

Instruments. The instruments involved were different for each campus, but the items for both instruments are responded to in a traditional Likert format. The form used at Campus 1 consisted of 18 usable items and the form used at campus 2 consisted of 27 items. Ten of the items on each form had a rough equivalent on the other form. Analyses were done using both the full set of items from each form, and for the 10 item subsets. Only the results for the comparable 10 item sets are reported here. Table 1 shows these item pairs for the two instruments.

Analysis. For the multivariate generalizability analyses, the methods suggested by Joe and Woodward (1976) were adopted. The two design configurations described above were analyzed using the BMD multivariate procedures to obtain the appropriate sums of squares and cross products matrices for the Joe and Woodward methods. Further analysis was done using specialized FORTRAN routines and several matrix subroutines from the IMSL Library. These specialized procedures were needed because the standard packages do not provide a way

of doing the multivariate analyses required for multivariate G studies.

All items which were negatively stated were rescaled to reflect a positive statement. All samples were analyzed with $n_s = 1$ and n_t(or n_c for courses within teachers) = 1 and with $n_s = 5$ and n_t (or n_c), = 2 where n_s is the number of students, n_t is the number of teachers and n_c is the number of courses to be used in the D study. The results reported here are for the latter analysis only (i. e., for $n_s = 5$ and n_t [or n_c] = 2).

Results

Analyses were done for each of the seven samples using the full complement of items and using only those items that were comparable across forms. The former of these analyses is not reported here for two reasons. First, the "noise" created in the data by non-comparable items across forms made interpretation nearly impossible. The specific pattern of "coefficients" for particular composites is somewhat dependent upon which items are included in the analysis, consequently comparability across the two forms seemed unreasonable. Second, the sample sizes that we are dealing with are generally small and we expect that we have capitalized on chance, and we expect that the more items we include, the more this will be true. This latter factor is addressed in a separate analysis reported after discussion of the analysis for the comparable ten item sets.

Instructors Within Courses

The results for the two samples with instructors nested within courses are given in Table 2, which presents, for the ten comparable items in the two instruments, the three composites with the highest generalizability coefficients. The three columns for each instrument contain the weights for

the three composites with the largest generalizability coefficients; the generalizability coefficient for each composite is given at the top of the column. If one considers the weights for all ten items, it is very difficult to interpret the composites, and therefore, this discussion will emphasize those items with the largest coefficients (positive or negative). Furthermore, it will be shown later that the estimates of coefficients are quite sensitive to the sampling of teachers, courses and students and, therefore, this discussion will concentrate on the general patterns among the coefficients and will not focus on the exact value of any weights in the composites.

For both of the analyses in Table 2, the first composite is dominated by the 14-7 and 13-17 item pairs, both of which have large positive coefficients. These two item pairs had to do with clarity of explanations and overall course evaluation. Therefore, if we focus only on the two items with the largest coefficients, the first composite for each of the instructor-within-course samples can be interpreted as indices of overall course quality with an emphasis on the clarity of explanations.

If we extend this interpretation to the third largest weight for the two analyses, item pair 8-16, which rates the method of instruction, is included. Item pair 8-16 has a fairly large, negative weight for the first composite in each of the two samples, and therefore complicates the interpretation. Considering the three item pairs with the largest coefficients, the first composite for the two samples in Table 2 appears to be a contrast of general course quality with method of instruction. Given that the objects of measurement for these analyses are courses rather than instructors, it is perhaps reasonable to subtract out some specific instructor characteristics, such as method of teaching. Following this pattern, item pair 11-6, treating

the instructor's knowledge of subject matter, also has substantial negative coefficients.

These results illustrate the problems that can arise in interpreting the composites generated by the Joe and Woodward procedures. It is often difficult to impose a reasonable interpretation on all of the weights in the composities. Furthermore, with small sample sizes and a large number of items an interpretation based on all of the details (the weights) generated by the analyses would not be statistically justified. Only a partial interpretation of the weights is practical. However, we have no rule for how much detail should be included in this partial interpretation, and the form of the interpretation depends on the amount of detail used in deriving the interpretations.

Although the interpretation of the second and third composites in the two analyses is less clear than the interpretation of the first, composite 2 from the Campus 1 form and composite 3 from the Campus 2 form exhibit certain similarities. Both have consistent high weights on the 12-14 and 9-9 combination and negative weights on the item pairs 6-11, 8-16, and 10-18. Again, this composite may be interpreted as a contrast of general course with specific instructor vaiables--in this case having to do with future relevance and plans for this topical area vs. focus on specific teaching behaviors. The evidence here is not strong, but given the relatively small sample size, does show some signs of promise.

Courses within Instructor

Tables 3 through 6 provide a comparison of the Campus 1 course within instructor results for each of the four Campus 2 course within instructor samples. Analyses for these samples focused upon differentiation of instruction/instructor. Examination of these tables indicates that the

course-within-instructor samples were even less consistent than the instructor-within-course samples discussed above. However, there are two fairly consistent patterns that can be seen across all five analyses.

First, each analysis contains a composite where the 6-11 item pair has the highest weight. Further comparisons among the other weights in these composites break down after this, but it may be speculated that a "learning" composite is important to the instructor discrimination. Second, across all samples there is a composite which has its highest weight on the 14-7 pair. Again, there is no consistent pattern for the other weights in these compositities, but the prominance of the 14-7 pair may suggest a dimension related to the instructor's ability to explain (or interpret) material. Although these speculations are highly tentative, they do have plausible interpretations and at least provide hypotheses for future examination of profile interpretations of CEP data.

Sampling Variability for Estimates of Composite Weights

As noted earlier, we had some concern about the effect of sampling teachers, courses and students on the estimated weights generated by the Joe and Woodward procedure. This concern was especially strong for some of the Campus 2 samples which had small sample sizes. Although the total number of students used in each analysis was fairly large, none of the samples had very many teachers and courses, and several of the Campus 2 samples had only five or six teachers. Since Smith (1978) and Shavelson and Webb (in press) have pointed out that relatively large sample sizes are typically needed for accurate estimates of variance components, these small sample sizes do not justify great confidence in the estimated weights.

Furthermore, there is reason to expect that the Joe and Woodward

estimation procedures will be even more susceptible to sampling variations than the estimates of variance components employed in univariate G studies. The Joe and Woodward procedure generates weighted composites with maximum generalizability. Since this procedure optimizes generalizability coefficients, given the data available, it will capitalize to some extent on the unique characteristics of particular samples employed (e.g., even though certain kinds of errors are uncorrelated in the universe of generalization, their correlation in the sample may be quite different from zero), and therefore is likely to generate different weights for independent replications of the analyses. The extent to which the procedures can capitalize on sampling variability will probably depend on the size of the sample (as the sample size increases, the deviation of sample statistics from population parameters should diminish) and the number of items (as the number of items increases, the opportunities to select weights that capitalize on the idosyncracies of the particular sample should increase). Since a fairly large number of items were used in the analyses reported in this paper and since the sample sizes for some sampling units were as small as five or six, the ability of the Joe and Woodward procedures to take advantage of random variations presents a potentially serious problem.

To provide some indication of the seriousness of this problem two of the data sets were split in half, and the Joe and Woodward procedure was applied separately to each half of the sample. By comparing the weights assigned to each item in the analyses for the two subsamples, some indication of the stability of the weights from sample to sample can be obtained.

For the first pair of split-halves analyses, the Campus 2 data set used for the analyses in Table 3 was split in half by assigning nine of the eighteen

students in each class to each of the two analyses. Therefore the two parallel analyses involved the same five teachers and the same sections for each teacher, but independent samples of students from the sections. This comparison keeps the teachers and sections constant and examines the effect on the composite weights of the sample of students, and provides an indication of the stability of the composite weights over samples of nine students (with five instructors and two sections).

The first three composites for these two analyses are reported in Table 7. It is clear from an examination of Table 7 that the specific weights assigned to different items are in many cases, quite different across the two analyses. Furthermore, the weights from the two analyses in Table 7 are also quite different from the weights found when the full sample was analyzed. However certain general patterns are consistent across the parallel analyses in Table 7. The first composite in each of the parallel analyses has a large positive weight for item 14 and a large negative weight for item 17, and this pattern also occurs in the second composite for the analyses of the full sample reported in Table 3.

Although a number of such similarities can be seen between the two analyses, there are also striking differences. For example, for the first of these split-halves analyses, item 8 has a substantial positive weight while, in the second analysis, item 8 has a substantial negative weight In general, the similarities between two parallel analyses are confined to the most general features of the composites. Since the weights for the full sample, as given in Table 3, are based on a sample of 18 students rather than 9 students, they can be expected to be somewhat more stable than the weig or the split-half analyses, but the results in Table 7 certainly do not inspire confidence in an interpretation of the details of the full analysis.

The second pair of parallel analyses is, if anything, more discouraging. Table 8 presents split-halves compaison of the course-within-instructor sample from Campus 1. Each of these two subsamples consisted of eight instructors, two courses per instructor and ten students from each course. Because the full sample was fairly large, these split-halves have sample sizes that are comparable to some of the analyses reported earlier in this paper. Again, there are certainly some strong similarities between the results for the two analyses; for example the first composite for the first analysis of Table 8 is quite similar to the third composite for the second analyses in Table 8. However, the differences between the two analyses are, even more compelling than the similarities. Certainly, the interpretation that could be given to the first analyses would be quite different from the interpretation for the second.

These two "split-halfs" analyses do not provide very extensive information about the stability of composites from the Joe and Woodward procedure; a different "split" of the samples might have produced very different results, or the agreement between the parallel analyses might have been better with a different design or slightly larger samples.

However, limited as they are, these split-halves analyses do suggest that for nested designs like those included in this study with about ten students per course, two courses per instructor and five to ten instructors, the application of the Woodward and Joe procedure does not produce very stable results. Although some general patterns in the composites seem to hold across parallel analyses, the specific values of the weights assigned to individual items are very variable from one analysis to the next.

If the Woodward and Joe procedures are to have any general utility, it is

necessary that the resulting composites be reasonably stable across different samples. In particular, the effects of various characteristics of the analyses, like sample sizes, complexity of design, number of variables (items in this paper), and intercorrelations among the variables, on the stability of the resulting composites must be better understood.

DISCUSSION

This paper has examined the possibility of applying the procedures suggested by Joe and Woodward to the analysis of course evaluation questionnaires. Several studies have examined the univariate generalizability of course evaluation questionnaires, and it was expected that a multivariate analysis would increase our understanding of the properties of such questionnaires.

One of the main results of the analyses reported in this paper has been an illustration of one potentially serious limitation of the Joe and Woodward procedure. In particular it was found that the weights assigned to the different items were highly variable across split-halves analyses. In a sense, this lack of generalizability for the composite weights was not very surprising. It is known that univariate generalizability studies require very large amounts of data (e.g., Smith, 1978) for the accurate estimation of variance components in complex designs. Because of its canonical algorithm, this is probably even more of a problem for the Joe and Woodward procedure than it is for univariate generalizability studies. Since the procedure finds composites that maximize generalizability coefficients for the data given, they

can capitalize on the idiosyncrasies of a particular sample, and since the opportunities for taking advantage of particular sample characteristics would increase with the number of variables, it is likely that the stability of the results across samples will be greatest when the number of variables is small. It may be that for designs of even moderate complexity, the number of variables must be very small (perhaps as small as 3 or 4) or the sample size must be very large to achieve stability.

With respect to course evaluations, due to small samples -- only tentative conclusions should be drawn. In our analysis, vague but viable patterns emerge which suggest possible profile discrimination and interpretation. As discussed earlier the tentativeness of this conclusion cannot be overstated, however.

References

Conger, A. J., & Lipshitz, R. Measures of reliability for profiles and test batteries. Psychometrika, 1973, 38, 411-428

Cronbach, L. J., Gleser, G., Nanda, N., & Rajaratnam, N. The dependability of behavioral measurements: The theory of generalizability for scores and profiles. New York: John Wiley & Sons, 1972.

Doyle, K. O. Student evaluation of instruction. Lexington, MA: Lexington Books, D. C. Heath & Company, 1975.

Gillmore, G. M., Kane, M. T., & Naccarato, R. W. The generalizability of of student ratings of instruction. Journal of Educational Measurement, 1978, 15, 1-15.

Joe, G. W., & Woodward, A. J. Some developments in multivariate generalizability. Psychometrika, 1976, 41, 205-218.

Kane, M. T., & Brennen, R. The generalizability of class means. Review of Educational Research, 1977, 47, 267-292.

Kane, M. T., Gillmore, G. M., & Crooks, T. J. Student evaluations of teaching: The generalizability of class means. Journal of Educational Measurement, 1976, 13, 171-184.

Kaufman, A. S. WISC-R research: Implications for interpretation. School Psychology Digest, 1978, 8, 5-27.

Shavelson, R. J. and Webb, N. M. Generalizability theory: 1973-1980. British Journal of Mathematical and Statistical Psychology, (in press).

Smith, P. L. The generalizability of student ratings of courses: Asking the right questions. Journal of Educational Measurement, 1979, 16, 77-87.

Smith P. L. Sampling errors of variance components in small sample generalizability studies. Journal of Educational Statistics, 1978, 3, 319-346.

Webb, N. M., & Shavelson, R. J. Multivariate generalizability of General Educational Development ratings. Journal of Educational Measurement, 1981, 1, 13-22.

Table 1

Item Content for Each of the Ten Item Pairs Used in Analysis*

Campus 2	ITEM NUMBER	Campus 1
Not much was gained by taking this course	6 - 11	Compare this course on the basis of learning principles and theories
The instructor encouraged the development of new view points and appreciations	7 - 2	The instructor encouraged students to express themselves freely and openly
I learn more when other methods are used	8 - 16	I would have preferred another method of teaching
The course material seemed worthwhile	9 - 9	Compare this course on its relevance to your course of study
The instructor was excellent	10 - 18	The instructor was excellent
The instructor demonstrated a thorough knowledge of the subject matter	11 - 6	The instructor demonstrated a thorough knowledge of the subject matter
I would rather not take another course from this instructor	12 - 14	I would take another course from this instructor
It was a very worthwhile course	13 - 17	The course was excellent
Some things were not explained very well	14 - 7	The instructor explained course material clearly
The instructor seemed to consider teaching a chore or routine activity	17 - 3	The instructor seemed enthusiastic about the subject matter.

* negatively stated items' scores were reversed for comparability.

Table 2

Multivariate Generalizability Coefficients and Composite Coefficients for Instructor-within-Course Designs (First Three from Each Campus Shown)

		Campus 1			Campus 2 (Sample 6)		
ITEM PAIR							
Campus 2	Campus 1	.82	.73	.66	.89	.78	.61
6	11	.353	-.380	.377	.054	-.447	-.572
7	2	-.198	.048	1.000	-.307	.217	.110
8	16	-.546	-.561	.263	-.436	-.308	-.358
9	9	.114	.318	.080	.413	.365	.667
10	18	.012	-.933	-.413	.304	-.244	-.242
11	6	-.502	.076	-.279	-.197	.676	-.427
12	14	-.338	1.000	-.031	-.044	.058	1.000
13	17	.594	.435	-.601	.811	.825	-.008
14	7	1.000	.026	.712	1.000	-.833	-.003
17	3	.054	-.129	-.735	-.064	1.000	.051

Table 3

Multivariate Generalizability Coefficients and Composite Coefficients for Course-within-Instructor Design; Campus 2 Sample 3 (First Three in Each Sample Shown)

		Campus 1			Campus 2 (Sample 3)		
ITEM PAIR							
Campus 2	Campus 1	.99	.92	.82	.69	.54	.37
6	11	1.000	.088	-.973	-.280	-.567	1.000
7	2	.538	.316	.059	.614	.689	.434
8	16	.251	-.528	.432	.344	-.240	.013
9	9	.003	-.305	.611	-.357	.110	-.828
10	18	-.454	.436	-.471	-.496	.149	-.873
11	6	.439	-.275	.066	-.190	.384	-.040
12	14	-.102	.569	-.190	.642	-.002	-.366
13	17	-.802	.712	1.000	.020	-.049	-.053
14	7	.894	1.000	-.290	-.235	1.000	.484
17	3	.442	-.266	.291	1.000	-.998	-.008

Table 4

Multivariate Generalizability Coefficients and Composite Coefficients for Course-within-Instructor Design; Campus 2 Sample 4 (First Three in Each Sample Shown)

ITEM PAIR		Campus 1			Campus 2 (Sample 4)		
Campus 2	Campus 1	.99	.92	.82	.92	.89	.74
6	11	1.000	.088	-.973	.294	-.718	1.000
7	2	.538	.316	.059	-.050	-.203	.887
8	16	.251	-.528	.432	-.249	-.007	-.173
9	9	.003	-.305	.611	-.619	1.000	-.121
10	18	-.454	.436	-.471	-.218	-.225	-.858
11	6	.439	-.275	.066	-.096	.132	.803
12	14	-.102	.569	-.190	-.170	-.198	-.488
13	17	-.802	.712	1.000	.119	.273	.020
14	7	.894	1.000	-.290	1.000	.021	-.880
17	3	.442	-.266	.291	-.311	.026	.045

Table 5

Multivariate Generalizability Coefficients and Composite Coefficients for Course-within-Instructor Design; Campus 2 Sample 5 (First Three in Each Sample Shown)

		Campus 1			Campus 2 (Sample 5)		
ITEM PAIR							
Campus 2	Campus 1	.99	.92	.82	.89	.82	.69
6	11	1.000	.088	-.973	1.000	-.648	-.083
7	2	.538	.316	.059	.330	.133	.125
8	16	.251	-.528	.432	-.274	-.554	-.048
9	9	.003	-.305	.611	-.496	1.000	.880
10	18	-.454	.436	-.471	.049	-.093	.098
11	6	.439	-.275	.066	-.121	-.098	-.003
12	14	-.102	.569	-.190	-.114	-.519	-.074
13	17	-.802	.712	1.000	.012	-.189	-.065
14	7	.894	1.000	-.290	.025	.793	1.000
17	3	.442	-.266	.291	-.512	-.310	-.204

Table 6

Multivariate Generalizability Coefficients and Composite Coefficients for Course-within-Instructor Design; Campus 2 Sample 7 (First Three in Each Sample Shown)

		Campus 1			Campus 2 (Sample 7)		
ITEM PAIR							
Campus 2	Campus 1	.99	.92	.82	.84	.73	.59
6	11	1.000	.088	-.973	1.000	-.486	-.181
7	2	.538	.316	.059	.313	-.148	-.001
8	16	.251	-.528	.432	-.046	-.178	-.816
9	9	.003	-.305	.611	-.133	.073	.795
10	18	-.454	.436	-.471	-.134	.225	.300
11	6	.439	-.275	.066	-.209	-.483	-.442
12	14	-.102	.569	-.190	-.602	-.475	-.058
13	17	-.802	.712	1.000	-.018	-.853	1.000
14	7	.894	1.000	-.290	-.854	1.000	.351
17	3	.442	-.266	.291	-.012	.319	-.243

Table 7

Campus 2, Sample 3 Split-half Analysis Where n_t = 2 and n_s = 9

(First Three Composites for Each Half Shown)

	Half 1			Half 2		
ITEM	.84	.80	.78	.76	.73	.71
6	-.720	-.594	.024	-.339	1.000	.269
7	-.090	1.000	.110	-.313	-.849	-.420
8	.455	.390	.183	-.538	-.599	-.275
9	.387	-.093	-.373	.013	-.136	-.144
10	-.008	.161	-.403	.139	-.236	-.383
11	-.432	.090	-.699	.016	-.227	.290
12	-.345	.283	.090	.041	-.071	-.271
13	-.010	.221	-.024	-.344	.095	.009
14	1.000	.468	.056	1.000	-.824	.138
17	-.621	-.430	1.000	-.689	.200	1.000

Table 8

Campus 1 Course-within-Instructor Split-half Analysis Where $n_c = 2$ and $n_s = 10$

(First Three Composites from Each Half Shown)

	Half 1			Half 2		
ITEM	.92	.89	.56	.88	.84	.59
11	-.571	.463	-.257	-.974	-.548	-.975
2	.348	1.000	-.135	.108	.048	.186
16	.530	-.918	-.453	-.055	.283	.471
9	-.180	-.696	.029	.504	-.625	-.566
18	-.122	-.009	.695	-.017	.673	-.984
6	-.182	.140	-.774	-.022	1.000	-.352
14	-.467	-.051	.666	-.123	-.467	-.379
17	1.000	-.135	.187	1.000	-.443	1.000
7	-.358	-.188	1.000	.167	-.621	-.639
3	-.030	-.047	-.158	.108	.049	.186

31

30-3-1